

## Patent Abstracts of Japan

PUBLICATION NUMBER : 2000250577  
PUBLICATION DATE : 14-09-00

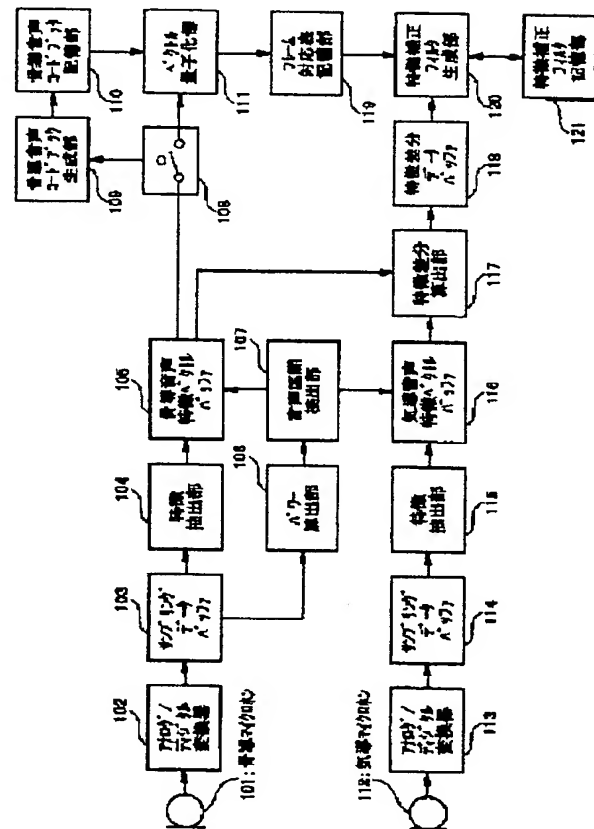
APPLICATION DATE : 24-02-99  
APPLICATION NUMBER : 11047262

APPLICANT : NIPPON TELEGR & TELEPH CORP  
<NTT>;

INVENTOR : SUZUKI YOSHITAKE;

INT.CL. : G10L 15/06 G10L 15/28 G10L 15/02

TITLE : VOICE RECOGNITION DEVICE AND  
LEARNING METHOD AND LEARNING  
DEVICE TO BE USED IN THE SAME  
DEVICE AND RECORDING MEDIUM  
ON WHICH THE SAME METHOD IS  
PROGRAMMED AND RECORDED



ABSTRACT : PROBLEM TO BE SOLVED: To use a bone-conduction microphone for voice recognition by improving the frequency characteristic of the microphone and also to enhance a recognition performance under a noisy environment.

SOLUTION: The feature vector of a voice (air-conduction sound) which is gathered in a second microphone 112 whose sound receiving sensitivity is secured in a frequency band broader than that of a first microphone 101 is estimated by inputting a voice input pattern (bone-conduction sound) gathered in the first microphone 101 to an analog-to-digital converter 102 and by selecting and outputting the feature vector while using a first code book 110 and by selecting a correction vector which corresponds to its index and is stored in a second code book 121 and by adding both vectors and by connecting them. Moreover, the estimated voice is used as a voice recognition object and voice parameters stored in a voice dictionary are successively applied to this object as patterns for reference and then both patterns of the voice pattern of the recognition object and the pattern for reference are collated in this device.

COPYRIGHT: (C)2000,JPO

EP 32747 (1)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号  
特開2000-250577  
(P2000-250577A)

(43) 公開日 平成12年9月14日 (2000.9.14)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テームト* (参考)
G 1 0 L 15/06		G 1 0 L 3/00	5 2 1 T 5 D 0 1 5
15/28			5 1 1
15/02			5 1 5 B
			5 2 1 U

審査請求 未請求 請求項の数10 O L (全 11 頁)

(21) 出願番号 特願平11-47262

(22) 出願日 平成11年2月24日 (1999.2.24)

(71) 出願人 000004226

日本電信電話株式会社

東京都千代田区大手町二丁目3番1号

(72) 発明者 井上 みづほ

東京都新宿区西新宿三丁目19番2号 日本  
電信電話株式会社内

(72) 発明者 鈴木 義武

東京都新宿区西新宿三丁目19番2号 日本  
電信電話株式会社内

(74) 代理人 100064908

弁理士 志賀 正武

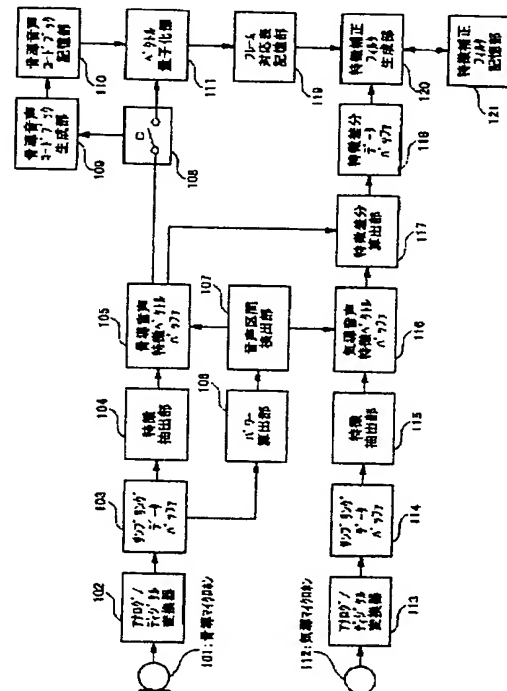
Fターム(参考) 5D015 DD01 FF04 FF05 GG00 HH04

(54) 【発明の名称】 音声認識装置及び同装置に使用される学習方法ならびに学習装置及び同方法がプログラムされ記録された記録媒体

(57) 【要約】

【課題】 骨導マイクロホンが持つ周波数特性の改善をはかり音声認識に用いる。合わせて騒音環境下での認識性能の向上をはかる。

【解決手段】 第一の受音器101で収録された音声入力パターン(骨導音)を第一のコードブック110を用いて特徴ベクトルを選択して出力し、そのインデックスに対応する第二のコードブック121に記憶されている補正ベクトルを選択し、前記両ベクトルを加算し接続することにより前記第一の受音器より広い周波数帯域で受音感度が確保される第二の受音器112で収録された音声(気導音)の特徴ベクトルを推定する。また、推定された音声を音声認識対象として用い、参照用パターンとして音声辞書213に格納された音声パラメータを順次与え両パターンの照合を行う。



## 【特許請求の範囲】

【請求項1】 音声信号を第一の受音器により受音し、受音した音声信号から、予め定めた時間長のフレーム毎に特徴ベクトルを抽出する手段と、抽出された特徴ベクトルを一時的に記憶する手段と、第一の受音器で予め受音した音声信号から抽出した代表的な有限個の特徴ベクトルを第一のセットとして記憶する手段と、前記第一の受音器で予め受音した音声信号の特徴ベクトルと前記第一の受音器よりも広い周波数帯域で受音感度が確保される第二の受音器で予め受音した音声信号の特徴ベクトルとの差分を用いて算出した代表的な有限個の補正ベクトルを第二のセットとして記憶する手段と、前記第一のセットに属する特徴ベクトルと前記第二のセットに属する各々の補正ベクトルを対応付ける手段と、前記第一のセットに属する各々の特徴ベクトルに対して、前記第一の受音器で受音した音声信号から抽出された特徴ベクトルの類似度を算出する手段と、類似度の最も高い特徴ベクトルを前記第一のセットの中から選択し、このベクトルに対応する前記第二のセットに属する補正ベクトルを抽出する手段と、前記第一の受音器で受音した音声信号から抽出された特徴ベクトルに対して前記抽出された補正ベクトルを加算して生成される特徴ベクトルをフレーム毎に算出する手段と、この特徴ベクトルの系列に対し、予め辞書として記憶された特徴ベクトル系列との間で類似度を照合する手段と、照合された中から最も類似度の高い辞書の情報を出力する手段とを備えることを特徴とする音声認識装置。

【請求項2】 入力される音声信号の特徴量から成るベクトルと前記第一のセットとして記憶されたベクトル間の距離で示される類似度を前記第一のセット中の少なくとも一部のベクトルについて計算して最小の特徴ベクトルを選択し、それに対応する補正ベクトルを加算することによって得られるベクトルから前記第二の受音器の特徴ベクトルを推定することを特徴とする請求項1記載の音声認識装置。

【請求項3】 前記第一の受音信号は発声者の頭骨を伝播する音響信号を受音する手段により生成され、前記第二の受音は、空気中を伝播する音響信号を受音する手段により生成されることを特徴とする請求項1または請求項2のいずれかに記載の音声認識装置。

【請求項4】 第一の受音器と、該第一の受音器より広い周波数帯域で受音感度が確保される第二の受音器とで収録された受音パターンを所定長の区間毎に切り出して特徴量を算出し、一方の受音器を介して抽出された特徴量をコードブックに記憶されたコードベクトルと比較することにより最も類似度の高い特徴ベクトルのインデックスを出力し、このインデックスに対応する他方のコードブックに記憶された補正ベクトルを出力することを特徴とする音声認識装置のコードブック学習方法。

【請求項5】 前記第一、第二の受音器で受音される信

号をフレーム毎同期をとって受音し、このうち、第一の受音器から第一のセットとして代表的な有限個の特徴ベクトルを抽出し、前記第一の受音器からの受音信号に対してフレーム毎に前記第一のセットに属する各々の特徴ベクトルの中で最も類似度の高い特徴ベクトルのインデックスを付与し、前記第一の受音器と同期する第二の受音器からの受音信号との差分を用い第二のセットとして代表的な有限個の補正ベクトルを抽出することにより、前記第一のセットに属する各々の特徴ベクトルと第二のセットに属する各々の補正ベクトルを対応付けることを特徴とする請求項4記載の音声認識装置のコードブック学習方法。

【請求項6】 前記第一の受音器を介して得られる音声信号から抽出した特徴ベクトルを適切な単語群の発声のある間蓄積したものをサンプルとし、各サンプル間の距離が小さいものをクラスターリングし、クラスター毎に特徴ベクトルの相加平均をとることにより代表的な特徴ベクトルを求め、これら代表的な特徴ベクトルを前記第一のセットに記録することを特徴とする請求項5記載の音声認識装置のコードブック学習方法。

【請求項7】 第一の受音器と、該第一の受音器より広い周波数帯域で受音感度が確保される第二の受音器で収録される受音パターンを所定の時間長の区間毎に切り出して特徴量を算出し、コードブックを参照することによって特徴ベクトルを抽出し、予め辞書として記憶されたコードベクトルとの間で類似度を照合することにより、照合された中から最も類似度の高い辞書情報を認識結果として出力する音声認識装置において、音声信号を第一の受音器により受音し、受音した音声信号から、所定の時間長のフレーム毎に特徴ベクトルを抽出する手段と、抽出された特徴ベクトルを一時的に記憶する手段と、前記第一の受音器で予め受音した音声信号から抽出した代表的な有限個の特徴ベクトルを第一のセットとして記憶する手段と、前記第一の受音器で予め受音した音声信号の特徴ベクトルと前記第二の受音器で予め受音した音声信号の特徴ベクトルとの差分を用い算出した代表的な有限個の補正ベクトルを第二のセットとして記憶する手段と、前記第一のセットに属する各々の特徴ベクトルと前記第二のセットに属する各々の補正ベクトルを対応付ける手段とを具備することを特徴とする音声認識装置のコードブック学習装置。

【請求項8】 前記第一の受音信号は発声者の頭骨を伝播する音響信号を受音する手段により生成され、前記第二の受音は、空気中を伝播する音響信号を受音する手段により生成されることを特徴とする請求項7記載の音声認識装置のコードブック学習装置。

【請求項9】 第一の受音器と、該第一の受音器より広い周波数帯域で受音感度が確保される第二の受音器で収録された受音パターンを所定時間長の区間毎に切り出して特徴量を算出し、コードブックを参照することによ

て特徴ベクトルを抽出する音声認識装置のコードブック学習装置に用いられ、前記第一の受音器と、第二の受音器で受音されたアナログ信号を適切なサンプリング周波数でデジタル信号に変換し、それぞれに用意されるサンプリングデータバッファに格納するステップと、それぞれのサンプリングデータバッファに格納されたデータをフレーム毎にデータの特徴量を算出するステップと、フレーム毎に算出されるパワーと事前に設定された閾値とを比較することにより、音声の開始フレーム及び終了フレームを決定するステップと、適切な単語の発声がある間、前記音声の開始フレームと終了フレームの情報に基づき当該範囲の特徴量のみを特徴ベクトルとして第一・第二のベクトルバッファに格納するステップと、前記第一のベクトルバッファに格納された特徴ベクトルから代表的な特徴ベクトルを生成し、第一のコードブックに格納するステップと、前記第一のベクトルバッファに格納された特徴ベクトルを、事前に生成され前記第一のコードブックに格納された特徴ベクトルに基づいてフレーム毎ベクトル量子化するステップと、前記第一のコードブックの特徴ベクトルの中で最も類似度の高い特徴ベクトルのインデックスを前記第一の特徴ベクトルに付与するステップと、前記第二のベクトルバッファに格納された特徴ベクトルと第一のベクトルバッファに格納された特徴ベクトルとの差分をフレーム毎算出し、その差分を特徴差分データバッファに格納するステップと、前記第一のベクトルバッファに格納された特徴ベクトルに対応する特徴差分データに対し、前記第一の特徴ベクトルに付与されたインデックスと同じインデックスを付与するステップと、前記特徴差分データに付与されたインデックス毎に特徴差分データをクラスタリングし、クラスタに含まれるデータを平均化することにより代表的な特徴補正ベクトルを生成して前記第二のコードブックに格納するステップが記録された記録媒体。

【請求項10】 受音器を介して収録される音声入力パターンを所定時間長の区間毎に切り出して特徴量を抽出し、コードブックを参照することにより特徴ベクトルを抽出し、予め辞書として記憶されたコードベクトルとの間で類似度を照合することにより、照合された中から最も類似度の高い辞書情報を出力する音声認識装置に用いられ、前記受音器で受音されたアナログ信号を適切なサンプリング周波数によりデジタル信号に変換しサンプリングデータバッファに格納するステップと、サンプリングデータバッファに1フレーム分の信号が格納される毎に信号パワーを算出し、フレーム毎に算出されるパワーを適切なフレーム数に渡って累積加算し、フレームあたりの平均値を計算して音声区間検出のための閾値とするステップと、フレーム毎に算出されるパワーと事前に設定された前記閾値とを比較することにより音声の開始フレーム及び終了フレームを決定するステップと、開始フレームと終了フレームの情報に基づき当該範囲の特徴

量をベクトルとして抽出し、特徴ベクトルバッファに格納するステップと、特徴ベクトルバッファに格納された特徴ベクトルを、事前に第一のコードブックに格納された特徴ベクトルに基づいてフレーム毎にベクトル量子化するステップと、前記第一のコードブックの特徴ベクトルの中で最も類似度の高い特徴ベクトルに付与されるインデックスとその特徴ベクトルを転送するステップと、事前に第二のコードブックに格納された特徴ベクトルから前記転送されたインデックスに相当する補正ベクトルを抽出し、この抽出された補正ベクトルを転送された特徴ベクトルに加算することにより前記第二のコードブックに格納される特徴ベクトルの推定を行うステップと、ここで得られる特徴ベクトルを入力パターンとし、参照パターンとして音声辞書に予め登録されてある、認識対象となる各単語の音声開始フレームから音声終了フレームまでの特徴パラメータを順次与え、単語毎両パターンの照合を行なうステップと、各参照パターン毎入力パターンとの距離値を算出し、全ての距離値の中で最小となる距離値に対応する参照パターン名を認識結果として出力するステップが記録された記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、騒音環境下でも認識性能の高い、音声認識装置及び同装置に使用される学習方法ならびに学習装置、及び同方法がプログラムされ記録される記録媒体に関する。

【0002】

【従来の技術】騒音環境下でも認識性能の高い音声認識装置を実現する方法が従来から提案されている。例えば、(1) 予め音声入力のない時点での騒音を測定し、音声入力時には予め測定した騒音の成分を減じる方法、(2) 音声入力用と騒音入力用の2個のマイクロホンを用いて騒音入力用マイクロホンへの入力信号の成分を音声入力用マイクロホンへの入力信号から減じる方法、(3) 予め測定した騒音で音声認識装置の辞書を学習する方法、(4) 受音器として骨導マイクロホンを使用する方法、などがあった。しかしながら上述した従来のいずれの方法によっても以下に示すように認識性能の向上のために問題を残している。具体的に、(1) に示す方法では、騒音の性質が常に一定の場合以外には効果が低下する。また、(2) に示す方法では、騒音の性質によらずある程度の効果はあるものの、マイクロホン同士が互いに近すぎ過ぎて設置されると、騒音用マイクロホンに音声が入り込み、騒音成分と共に音声成分の一部までも減じられてしまう。逆にマイクロホン同士が離れ過ぎて配置されると、両者に入力される騒音の性質が異なってしまうため、騒音成分の正確な減算ができない。更に、複数のマイクロホンを設置するために装置規模が大きくなったり、発声者の位置が制限されるといった諸々の問題を有していた。また、(3) に示す方法では、学習時

と認識時における騒音の性質が異なると効果が低下する。(4)に示す方法では、原理的に騒音の影響を受け難い利点はあるが、受信した音声の周波数帯域が狭いため、音声情報が欠落するといった問題を有していた。

#### 【0003】

【発明が解決しようとする課題】一方、音声入力用マイクロホンが持つマイクロホン特性の差や、音声入力から音声認識に至る伝送路特性の差を補正する方法としてフィルタを用いる方法が提案されている。そこで、音声の周波数帯域が狭い骨導マイクロホン(広い周波数帯域に渡って受信できる気導マイクロホンに比べ、受信できる周波数帯域が制限されるが空气中を伝播する騒音成分の影響が少ない)の特性を、周波数帯域が広いマイクロホン特性に補正するフィルタを用いる方法もあるが実用に十分な効果を得るには至っていないのが現状である。本発明は上記事情に基づいてなされたものであり、騒音の影響を受け難い骨導マイクロホンを受音器として用い、周波数特性を気導マイクロホンが持つ特性に近づけることにより、騒音環境下における音声認識性能の向上を、発声者の位置を制限することなく、装置規模を増大させることなく、かつ、従来の音声認識装置にも容易に適用できる、音声認識装置及び同装置が使用する学習方法ならびに学習装置、及び同方法がプログラムされ記録される記録媒体を提供することを目的とする。

#### 【0004】

【課題を解決するための手段】本発明の音声認識装置は、音声信号を第一の受音器により受信し、受信した音声信号から、予め定めた時間長のフレーム毎に特徴ベクトルを抽出する手段と、抽出された特徴ベクトルを一時的に記憶する手段と、第一の受音器で予め受信した音声信号から抽出した代表的な有限個の特徴ベクトルを第一のセットとして記憶する手段と、前記第一の受音器で予め受信した音声信号の特徴ベクトルと前記第一の受音器よりも広い周波数帯域で受信感度が確保される第二の受音器で予め受信した音声信号の特徴ベクトルとの差分を用いて算出した代表的な有限個の補正ベクトルを第二のセットとして記憶する手段と、前記第一のセットに属する特徴ベクトルと前記第二のセットに属する各々の補正ベクトルを対応付ける手段と、前記第一のセットに属する各々の特徴ベクトルに対して、前記第一の受音器で受信した音声信号から抽出された特徴ベクトルの類似度を算出する手段と、類似度の最も高い特徴ベクトルを前記第一のセットの中から選択し、このベクトルに対応する前記第二のセットに属する補正ベクトルを抽出する手段と、前記第一の受音器で受信した音声信号から抽出された特徴ベクトルに対して前記抽出された補正ベクトルを加算して生成される特徴ベクトルをフレーム毎に算出する手段と、この特徴ベクトルの系列に対し、予め辞書として記憶された特徴ベクトル系列との間で類似度を照合する手段と、照合された中から最も類似度の高い辞書の

情報を出力する手段とを備えることを特徴とする。

【0005】本発明の音声認識装置のコードブック学習方法は、第一の受音器と、該第一の受音器より広い周波数帯域で受信感度が確保される第二の受音器とで収録された受信パターンを所定長の区間毎に切り出して特徴量を算出し、一方の受音器を介して抽出された特徴量をコードブックに記憶されたコードベクトルと比較することにより最も類似度の高い特徴ベクトルのインデックスを出力し、このインデックスに対応する他方のコードブックに記憶された補正ベクトルを出力することを特徴とする。

【0006】本発明の音声認識装置のコードブック学習装置は、第一の受音器と、該第一の受音器より広い周波数帯域で受信感度が確保される第二の受音器で収録される受信パターンを所定の時間長の区間毎に切り出して特徴量を算出し、コードブックを参照することによって特徴ベクトルを抽出し、予め辞書として記憶されたコードベクトルとの間で類似度を照合することにより、照合された中から最も類似度の高い辞書情報を認識結果として出力する音声認識装置において、音声信号を第一の受音器により受信し、受信した音声信号から、所定の時間長のフレーム毎に特徴ベクトルを抽出する手段と、抽出された特徴ベクトルを一時的に記憶する手段と、前記第一の受音器で予め受信した音声信号から抽出した代表的な有限個の特徴ベクトルを第一のセットとして記憶する手段と、前記第一の受音器で予め受信した音声信号の特徴ベクトルと前記第二の受音器で予め受信した音声信号の特徴ベクトルとの差分を用いて算出した代表的な有限個の補正ベクトルを第二のセットとして記憶する手段と、前記第一のセットに属する各々の特徴ベクトルと前記第二のセットに属する各々の補正ベクトルを対応付ける手段とを具備することを特徴とする。

【0007】本発明の記録媒体は、第一の受音器と、該第一の受音器より広い周波数帯域で受信感度が確保される第二の受音器で収録された受信パターンを所定時間長の区間毎に切り出して特徴量を算出し、コードブックを参照することによって特徴ベクトルを抽出する音声認識装置のコードブック学習装置に用いられ、前記第一の受音器と、第二の受音器で受信されたアナログ信号を適切なサンプリング周波数でディジタル信号に変換し、それぞれに用意されるサンプリングデータバッファに格納するステップと、それぞれのサンプリングデータバッファに格納されたデータをフレーム毎にデータの特徴量を算出するステップと、フレーム毎に算出されるパワーと事前に設定された閾値とを比較することにより、音声の開始フレーム及び終了フレームを決定するステップと、適切な単語の発声がある間、前記音声の開始フレームと終了フレームの情報に基づき当該範囲の特徴量のみを特徴ベクトルとして第一・第二のベクトルバッファに格納するステップと、前記第一のベクトルバッファに格納され

た特徴ベクトルから代表的な特徴ベクトルを生成し、第一のコードブックに格納するステップと、前記第一のベクトルバッファに格納された特徴ベクトルを、事前に生成され前記第一のコードブックに格納された特徴ベクトルに基づいてフレーム毎ベクトル量子化するステップと、前記第一のコードブックの特徴ベクトルの中で最も類似度の高い特徴ベクトルのインデックスを前記第一の特徴ベクトルに付与するステップと、前記第二のベクトルバッファに格納された特徴ベクトルと第一のベクトルバッファに格納された特徴ベクトルとの差分をフレーム毎算出し、その差分を特徴差分データバッファに格納するステップと、前記第一のベクトルバッファに格納された特徴ベクトルに対応する特徴差分データに対し、前記第一の特徴ベクトルに付与されたインデックスと同じインデックスを付与するステップと、前記特徴差分データに付与されたインデックス毎に特徴差分データをクラスタリングし、クラスタに含まれるデータを平均化することにより代表的な特徴補正ベクトルを生成して前記第二のコードブックに格納するステップが記録されることを特徴とする。

【0008】また、受音器を介して収録される音声入力パターンを所定時間長の区間毎に切り出して特徴量を抽出し、コードブックを参照することにより特徴ベクトルを抽出し、予め辞書として記憶されたコードベクトルとの間で類似度を照合することにより、照合された中から最も類似度の高い辞書情報を出力する音声認識装置に用いられ、前記受音器で受音されたアナログ信号を適切なサンプリング周波数によりデジタル信号に変換しサンプリングデータバッファに格納するステップと、サンプリングデータバッファに1フレーム分の信号が格納される毎に信号パワーを算出し、フレーム毎に算出されるパワーを適切なフレーム数に渡って累積加算し、フレームあたりの平均値を計算して音声区間検出のための閾値とするステップと、フレーム毎に算出されるパワーと事前に設定された前記閾値とを比較することにより音声の開始フレーム及び終了フレームを決定するステップと、開始フレームと終了フレームの情報に基づき当該範囲の特徴量をベクトルとして抽出し、特徴ベクトルバッファに格納するステップと、特徴ベクトルバッファに格納された特徴ベクトルを、事前に第一のコードブックに格納された特徴ベクトルに基づいてフレーム毎にベクトル量子化するステップと、前記第一のコードブックの特徴ベクトルの中で最も類似度が高い特徴ベクトルに付与されるインデックスとその特徴ベクトルを転送するステップと、事前に第二のコードブックに格納された特徴ベクトルから前記転送されたインデックスに相当する補正ベクトルを抽出し、この抽出された補正ベクトルを転送された特徴ベクトルに加算することにより前記第二のコードブックに格納される特徴ベクトルの推定を行うステップと、ここで得られる特徴ベクトルを入力パターンとし、

参照パターンとして音声辞書に予め登録されてある、認識対象となる各単語の音声開始フレームから音声終了フレームまでの特徴パラメータを順次与え、単語毎両パターンの照合を行なうステップと、各参照パターン毎入力パターンとの距離値を算出し、全ての距離値の中で最小となる距離値に対応する参照パターン名を認識結果として出力するステップが記録されることも特徴とする。

【0009】このことにより、騒音の影響を受け難い骨導マイクロホンを受音器として用い、周波数特性を補正するフィルタを用いることによってその周波数特性を気導音声に近づけることができ、発声者の位置を制限することなく、装置規模を増大させることなくして騒音環境下における音声認識性能の向上がはかれ、また、従来からある音声認識装置にも容易に適用できる。

【0010】

【発明の実施の形態】図1は、本発明における音声認識装置のコードブック学習装置の実施形態を示すブロック図である。図において、101は骨導マイクロホン、112は気導マイクロホンである。気導マイクロホン112は、広い周波数帯域に渡って感度が良いことで知られている。音声認識に必要な8キロヘルツ乃至12キロヘルツの帯域の音響信号を良好に受信できる。一方、音声と同一周波数帯域の騒音信号も音声信号と区別なく受信できるため、高騒音下での音声区間検出が困難になるという欠点を持つ。骨導マイクロホン101は、加速度ピックアップを用いているため周波数帯域が狭く、音声の高域周波数成分の減衰が大きい。そのため、単独で音声認識に用いると性能が低下する反面、外部からの空気中を伝播する騒音成分の影響が小さいという利点を持つ。尚、本発明では、骨導マイクロホン101、気導マイクロホン112によって収録される音声入力パターンは、一定時間区間毎に切り出され、それぞれに用意されるコードブックと比較して最も類似度の高いコードベクトルを取り出すベクトル量子化手法(VQ: vector quantization)によって音声入力パターンが表現されるものとして以降説明する。コードブックに保存されるベクトルをパターン空間内の重心(セントロイド: centroid)とする方法の代表例は、LBG法(Linde, Y. Buzo, A. and Gray, R. M.: "An Algorithm for vector quantizer design" IEEE Trans. Commun., COM-28, 1, 84-95(1980))として公知である。

【0011】前記骨導マイクロホン101、気導マイクロホン112により受信された音声信号は、アナログデジタル変換器(以下、単にA/D変換器という)102、113にそれぞれ供給され、A/D変換器102、113によって生成されるデジタル信号は、それぞれサンプリングデータバッファ103、114に供給される。サンプリングデータバッファ103、114出力はそれぞれ特徴抽出部104、特徴抽出部115に供給される。特徴抽出部104、115以降説明されるブロッ



クは全てソフトウェアによって実現されるものであり、ここでは機能ブロックとして表現してある。サンプリングデータバッファ103出力は更に、パワー算出部106、音声区間検出部107の経路により、骨導音声特徴バッファ105、気導音声特徴ベクトルバッファ116に供給される。108は切替スイッチである。切替スイッチ108は、骨導音声特徴ベクトルバッファ105出力を、骨導音声コードブック生成部109または、ベクトル量子化部111に接続する。110は骨導音声コードブック生成部109により生成される骨導音声コードを記憶する骨導音声コードブック記憶部である。一方、骨導音声特徴ベクトルバッファ105出力は切替スイッチ108の他に特徴差分算出部117に供給される。特徴差分算出部117で算出される骨導音声特徴ベクトルバッファ105と気導音声特徴ベクトルバッファ116の特徴ベクトル差分は特徴差分データバッファ118に供給される。特徴差分データバッファ118に供給されるデータは特徴補正フィルタ生成部120に供給される。特徴補正フィルタ生成部120に供給される差分データは後述するロジックにより特徴補正ベクトル成分として生成され、特徴補正フィルタ記憶部121ならびにフレーム対応表記憶部119に供給される。フレーム対応表記憶部119は、後述するロジックに従い特徴補正フィルタ生成部120により生成される補正ベクトルをベクトル量子化部111によって出力される骨導音声における特徴ベクトルの量子化出力に反映させる。

【0012】図3は図1に示す音声認識装置のコードブック学習装置の動作手順を説明するために引用したフローチャートであり、具体的にその手順は、本発明の学習装置の中にプログラムされ記録される。図示せぬCPUがこれを読み出し実行することにより以下に示す手順が実行される。以下、図1に示す音声認識装置のコードブック学習装置の動作について図4に示すフローチャートを参照しながら詳細に説明する。動作は機能的に大別すると、「骨導音声コードブック生成」と「特徴補正ベクトル生成」に区分される。まず、「骨導音声コードブック生成」動作から説明する。骨導マイクロホン101と気導マイクロホン112で受音されたアナログ信号は、それぞれA/D変換器102、113において適切なサンプリング周波数でデジタル信号に変換され（ステップS31、S32）、サンプリングデータバッファ103、114にそれぞれ逐次格納される（ステップS33）。ここで、適切なサンプリング周波数とは、音声認識処理に必要な音声の特徴を損なわない周波数であり、通常は、8キロヘルツから12キロヘルツに設定される。特徴抽出部104、115では、サンプリングデータバッファ103、114に20乃至30ミリ秒分のデータが格納される毎にデータの特徴量を算出する。すなわち、フレーム毎にデータの特徴抽出を行う（ステップS35）。一方、骨導マイクロホン101のサンプリ

ングデータからは、フレーム毎にパワー算出部106において信号パワーが算出され、音声区間検出部107に送られる。音声区間検出部107では、フレーム毎に算出されるパワーと事前に設定された閾値とを比較することにより、音声の開始フレーム及び終了フレームが決定される（ステップS34）。骨導マイクロホン101と気導マイクロホン112における各々のフレーム毎の特徴量は、検出された開始フレームと終了フレームの情報に基づき、当該範囲の特徴量のみ、ベクトルとして各々の特徴ベクトルバッファ105、116に格納される（ステップS37）。この操作は適切な単語の発声がある間繰り返される（ステップS36）。

【0013】ここで、適切な単語の発声がある間とは、全ての音韻の発声頻度の偏りが少ない単語群が出力されている間のことである。また、ここで注意すべきことは、骨導音声を用いて特徴量を算出することと、パワーを算出するフレームと気導音声を用いて特徴量を算出するフレームとの同期がとられることである。具体的に、パワーを算出するフレームと気導音声を用いて特徴量を算出するフレームとの同期は、A/D変換器102、113双方のサンプリングクロックの同期をとることにより容易に実現することができる。骨導音声コードブック生成時、切替スイッチ108は、骨導音声特徴ベクトルバッファ105を骨導音声コードブック生成部109に接続される（ステップS38）。骨導音声コードブック生成部109では、骨導音声特徴ベクトルバッファ105に格納された特徴ベクトルから代表的な特徴ベクトルを生成し（ステップS39）、骨導音声コードブック110に格納する（ステップS40）。尚、上述した代表的な特徴ベクトルは、骨導マイクロホン101の音声信号を介して抽出した特徴量を適切な単語群の発声がある間蓄積したものをサンプルとし、各サンプル間の距離が小さいもの同士をクラスタリングし、各クラス毎に特徴量の相加平均をとることにより求められる。ここで求められた代表ベクトルを記録して骨導音声コードブック（骨導音声コードブック記憶部110）とするものである。

【0014】次に、「特徴補正ベクトル生成」動作について説明する。特徴補正ベクトル生成時、切替スイッチ108はベクトル量子化部111に接続される（ステップS48）。ベクトル量子化部111では、骨導音声特徴ベクトルバッファ105に格納された特徴ベクトルを、事前に生成され骨導音声コードブック記憶部110に格納された特徴ベクトルに基づいてフレーム毎ベクトル量子化し（ステップS41）、骨導音声コードブックの特徴ベクトルの中で最も類似度の高い特徴ベクトルのインデックス（番号）を骨導音声特徴ベクトルに付与する（ステップS42）。一方、特徴差分算出部117では、骨導音声特徴ベクトルバッファ105に格納された特徴ベクトルと気導音声特徴ベクトルバッファ116に

格納された特徴ベクトルとの差分をフレーム毎算出し、特徴差分データバッファ118に逐次格納する。フレーム対応表記記憶部117では、骨導音声特徴ベクトルに対応する特徴差分データバッファ118に格納された特徴差分データに対し、骨導音声特徴ベクトルに付与した番号と同じ番号を付与する。特徴補正フィルタ生成部120では、特徴差分データに付与された番号毎に特徴差分データをクラスタリングし、クラスタに含まれるデータを平均化することにより代表的な特徴補正ベクトルを生成し(ステップS53)、特徴補正フィルタ記憶部121に格納する(ステップS54)。上述した骨導音声コードブックと特徴補正フィルタは番号毎対応がとれたものになっている。尚、上述した特徴補正ベクトルの生成は、骨導マイクロホン101からの音声信号と同期をとって得られた気導マイクロホン112からの音声信号が減算された残差信号(特徴差分算出部117)から抽出した特徴量を抽出することによりなされる。

【0015】図2は、本発明における音声認識装置の実施形態を示すブロック図である。図において、201は骨導マイクロホン、202はA/D変換器、203はサンプリングデータバッファである。また、204は特徴抽出部、205は特徴ベクトルバッファ、206はパワー算出部、207は音声区間検出部、208はベクトル量子化部、209は骨導音声コードブック記憶部、210は気導音声推定部、211は特徴補正フィルタ記憶部、212はパターン照合部、213は音声辞書記憶部、214は認識結果出力部であり、いずれもソフトウェアにより実現されるため、機能ブロックとして示した。機能等動作手順について詳細は後述する。

【0016】図4は、図2に示す音声認識装置の動作手順を説明するために引用したフローチャートであり、

(a)に騒音測定動作、(b)に音声認識動作についての手順が示されている。具体的にその手順は、本発明の音声認識装置の中にプログラムされ記録される。図示せぬCPUがこれを読み出し実行することにより以下に示す手順が実行されるものである。以下、図2に示す音声認識装置の動作について図4(a)(b)に示すフローチャートを参照しながら詳細に説明する。本発明の音声認識装置の動作は、機能的に大別すると、音声が入力の状態での騒音レベルを測定して、音声区間の閾値を決定する「騒音測定」と、発声された音声パターンと既登録の音声辞書中の音声パターンとを照合し、その結果を出力する「音声認識」に区分される。まず、「騒音測定」動作から説明する。骨導マイクロホン201で受信されたアナログ信号はA/D変換器202でデジタル信号に変換され、サンプリングデータバッファ203に逐次格納される(ステップS51、S52)。サンプリングデータバッファ203に1フレーム分の信号が格納される毎に、パワー算出部206において信号パワーが算出される(ステップS53)。そして、フレーム毎に

算出されたパワーは音声区間検出部207に入力される。音声区間検出部207では、フレーム毎に算出されるパワーを適切なフレーム数にわたって累積加算し、更にフレームあたりの平均値を計算する。ここで、適切なフレーム数とは、通常4乃至16程度である。算出された平均パワーに適切な定数を加算することにより、音声区間検出のための閾値とする(ステップS54)。

【0017】次に、「音声認識」動作について説明する。上述した「騒音測定」が終了した時点で音声入力が可能となる。まず、骨導マイクロホン201で受信された信号は、A/D変換器202でデジタル信号に変換され、サンプリングデータバッファ203に逐次格納される(ステップS61、S62)。1フレーム毎、特徴抽出部204でデータの特徴量が算出され、同時にパワー算出部206でその信号のパワーが算出され(ステップS63)音声区間検出部207に送られる。音声区間検出部207では、フレーム毎に算出されるパワーと事前に設定された閾値を比較することにより、音声の開始フレーム及び終了フレームが決定される(ステップS64)。ここで検出された開始フレームと終了フレーム情報に基づき、当該範囲の特徴量が、ベクトルとして特徴ベクトルバッファ205に格納される(ステップS65)。ベクトル量子化部208では、特徴ベクトルバッファ205に格納された特徴ベクトルを、事前に骨導音声コードブック記憶部209に格納された特徴ベクトルに基づいて、フレーム毎ベクトル量子化し(ステップS66)、骨導音声コードブックの特徴ベクトルの中で最も類似度が高い特徴ベクトルの番号と特徴ベクトルバッファ205に格納された特徴ベクトルが気導音声推定部210に転送される(ステップS67)。気導音声推定部210では、事前に特徴補正フィルタ記憶部211に格納された特徴ベクトルから転送された番号にあたる補正ベクトルを抽出し、この抽出された補正ベクトルを転送された特徴ベクトルに加算することによって気導音声特徴ベクトルへの推定が行なわれる(ステップS68)。尚、上述した類似度とは、骨導音声による特徴量から成るベクトルと骨導コードブックに記憶されたベクトルとの間の距離である。すなわち、両ベクトルの差の自乗値を要素毎に加算して得られた値である。ここでは、この類似度を骨導コードブック中の各ベクトル毎に計算し、最も小さくなるベクトルを選択する。このようにして選択され骨導コードブックから得られるベクトルと補正ベクトルを加算することによって気導音声特徴ベクトルを推定している。

【0018】一方、音声辞書記憶部213には、認識対象となる各単語の音声開始フレームから音声終了フレームまでの特徴パラメータが登録されている。従って、パターン照合部212への入力パターンとして、気導音声推定部210で推定された気導音声特徴ベクトルを与え、参照用パターンとして音声辞書記憶部213に格納



された音声パラメータを単語毎順次与えることにより、両パターンとの照合（ステップS69）を行なうことができる。その結果、入力パターンと参照パターンの照合結果が、距離値により出力される。距離値が大きいほど両パターンの相違度が大きいことを意味する。各参照パターン毎に入力パターンとの距離値を算出し（ステップS70）、全ての距離値の中で最小となる距離値に対応する参照パターン名が認識結果となり、認識結果表示部214に表示される（ステップS71）。尚、図3、図4に示すフローチャートは、それぞれ、学習装置、音声認識装置が持つ記憶装置（図示せず）中に固定的に書き込まれるか、あるいは半導体記憶装置、フロッピーディスクやハードディスク等の磁気記録装置、CD-ROM等にプログラムとして書き込まれて頒布されるものであり、必要に応じて装置内部の記憶装置に取込まれることにより機能するものである。

【0019】出願人は、上述した本発明実施形態による効果を確認するために以下に示す音声認識実験を行なった。図5は、図2に示す音声認識装置において、コードブックサイズに応じて増加される特徴補正フィルタ分割数毎の単語音声認識率の推移をグラフ表示したものである。グラフはX軸にフィルタ分割数を、Y軸に音声認識率を目盛り両者の関係を示している。単語は、電子協100都市名の最初の20都市を選び、発声は予め定めた騒音環境下で男女各2名が2回ずつ発声し、そのときの平均を認識率として示している。実験の結果、単一の特徴補正フィルタを用いた場合（フィルタ分割数1）、認識率が52.5%であったのに対し、コードブックサイズに応じてフィルタ分割数を増加させた場合、フィルタ分割数が64で認識率80%、128で82.2%、256で88.5%となった。このことにより無騒音下における本発明の効果が確認された。次に、騒音下での本発明の効果につき述べる。騒音環境下で、気導音声を入力とする従来の音声認識装置と、図2に示す本発明の音声認識装置との単語音声認識性能比較を行った。尚、本発明の音声認識装置において使用される補正フィルタの分割数は256とした。また、単語は、電子協100都市名の最初の20都市を選び、予め定めた環境下で2回ずつ発声した。結果、男女各2名の音声認識率の平均値は、騒音が64dBのピンクノイズをラウドスピーカより発生させた場合、従来装置が42.5%であったのに対し、本発明装置では79%であった。また、騒音が環境騒音（非正常騒音、環境騒音1は道路沿い、最大80dB、最小55dB、平均66dB、環境騒音2はショッピングモール、最大70dB、最小60dB、平均64dB）をラウドスピーカより発生させた場合、従来装置はいずれも認識率0%と全く認識できなかったのに対し、本発明装置では、それぞれ、認識率82.1%、80.4%であった。この性能比較により騒音環境下における本発明効果が確認された。

【0020】尚、出願人は、同日付で、騒音の影響を受け難い骨導マイクロホンと、周波数帯域の広い気導マイクロホンを併用し、骨導音声から気導音声への特徴ベクトルのマッピングを用いることによって騒音環境下における音声認識性能の向上をはかった。音声認識装置ならびに同装置における音声学習方法ならびに装置及び同方法がプログラムされ記録される記録媒体を出願してある。これに対し、本発明は、騒音の影響を受け難い骨導マイクロホンを受音器として用い、かつ、周波数を補正するフィルタを付加することによって周波数特性を気導音声に近づけることにより騒音下での音声認識性能の向上をはかったものである。このため、本発明の学習装置においては、骨導音声特徴ベクトルバッファ105に格納された特徴ベクトルと気導音声特徴ベクトルバッファ116に格納された特徴ベクトルとのベクトル差分をフレーム毎に算出する特徴差分算出部117と、これを記憶する特徴差分データバッファ118が付加され、更に、ここで得られる特徴差分データに付与された番号（フレーム対応記憶部で骨導音声特徴ベクトルに対応する特徴差分データ差分データバッファに格納された特徴差分データに対し骨導音声特徴ベクトルに付与された番号と同じ番号を付与している）毎に特徴差分データをクラスタリングし、クラスタに含まれるデータを平均化して代表的な特徴補正フィルタを生成するロジックを持つ特徴補正フィルタ生成部120が付加されている。また、本発明の音声認識装置においては、特徴補正フィルタ記憶部211が付加され、ここでは、気導音声推定部210で事前に特徴補正フィルタ記憶部211に記憶された補正フィルタから、転送された番号（ベクトル量子化部208から骨動音声コードブックの特徴ベクトル中で最も類似度の高い特徴ベクトルの番号と特徴ベクトルバッファに格納された特徴ベクトルが転送）にあたる補正フィルタを抽出し、転送された特徴ベクトルに抽出された補正フィルタを加算し、気導音声特徴ベクトルの推定を行うロジックを持つ。このことにより、骨導マイクロホンを受音器とし、周波数特性を補正するフィルタを介して骨導マイクロホンが持つ周波数特性の改善をはかり、気導音声に近づけると共に騒音下での音声認識性能の向上を実現するものである。

【0021】

【発明の効果】以上説明のように本発明は、第一の受音器で収録された音声入力パターン（骨導音）を第一のコードブックを用いて特徴ベクトルを選択して出力し、そのインデックスに対応する第二のコードブックに記憶されている補正ベクトルを選択し、前記両ベクトルを加算し接続することにより前記第一の受音器より広い周波数帯域で受音感度が確保される第二の受音器で収録された音声（気導音）の特徴ベクトルを推定し、推定された音声を音声認識対象として用いるものであり、このことにより、骨導マイクロホンが従来から特徴として持つ、外

部からの空气中を伝播する騒音成分の影響が小さいという利点を生かしながら、加速度ピックアップを使用しているため、周波数帯域が狭く、音声の高周波成分の減衰が大きいといった周波数特性の改善がはかれ、従って、この骨導マイクロホンを音声認識のための音声収録マイクロホンとして単独で 사용할 ことができる。また、騒音環境下における音声認識性能の向上をはかることができ、発声者の位置を制限することなく、装置規模を増大させることなく実現でき、また、従来からある音声認識装置にも容易に適用できるものである。

【図面の簡単な説明】

【図1】 本発明の学習装置の実施形態を示すブロック図である。

【図2】 本発明の音声認識装置の実施形態を示すブロック図である。

【図3】 図1に示す本発明実施形態の動作を説明するために引用したフローチャートである。

【図4】 図2に示す本発明実施形態の動作を説明するために引用したフローチャートである。

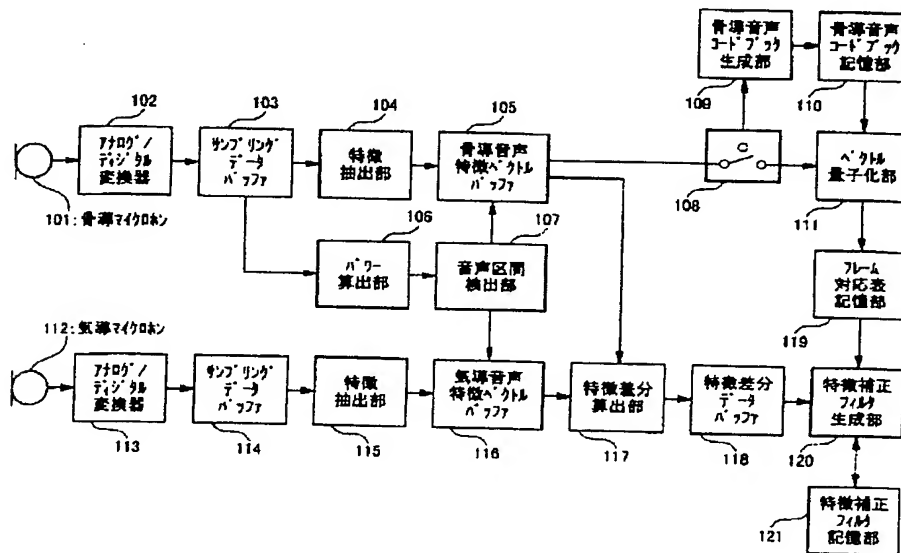
【図5】 本発明実施形態の効果を説明するために引用

したグラフである。

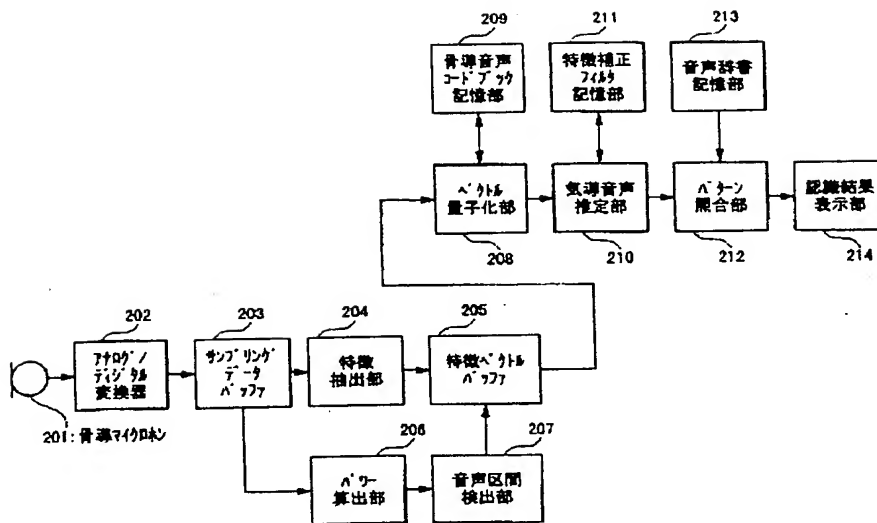
【符号の説明】

101、201…骨導マイクロホン（第一の受音器）、102、113、202…アナログデジタル変換器（A/D変換器）、103、114、203…サンプリングデータバッファ、104、115、204…特徴抽出部、105…骨導音声特徴ベクトルバッファ、106、206…パワー算出部、107、207…音声区間検出部、108…切替スイッチ、109…骨導音声コードブック生成部、110、209…骨導音声コードブック記憶部（第一のコードブック）、111、208…ベクトル量子化部、112…気導マイクロホン（第二の受音器）、116…気導音声特徴ベクトルバッファ、117…特徴差分算出部、118…特徴差分データバッファ、119…フレーム対応表記憶部、120…特徴補正フィルタ生成部、121、211…特徴補正フィルタ記憶部（第二のコードブック）、205…特徴ベクトルバッファ、210…気導音声推定部、212…パターン照合部、213…音声辞書記憶部（音声辞書）、214…認識結果表示部

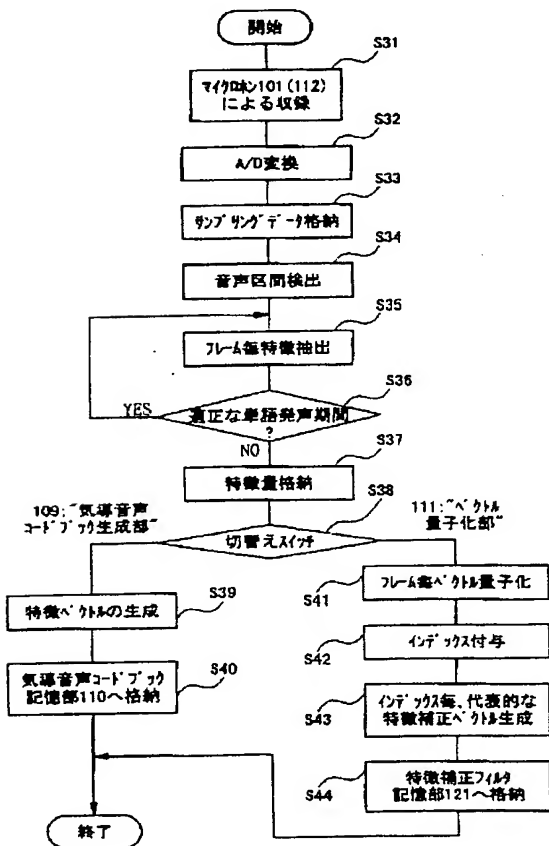
【図1】



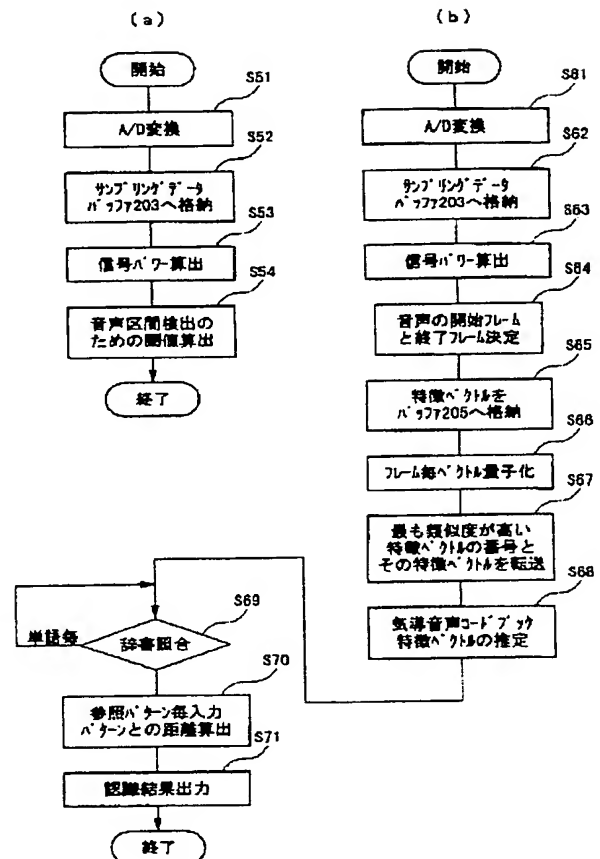
【図2】



【図3】



【図4】



(特1) 000-250577 (P2000-P0077)

【図5】

